OFFICE OF THE DATA PROTECTION COMMISSIONER

Guidance Note on

Privacy-Enhancing Technologies

July 2026.

CONTENTS

DEFINITION OF TERMS

"Act" means the Data Protection Act, Cap. 411C.

"Anonymisation" means the processing of personal data in a manner that renders the re-identification of the data subject not reasonably likely, directly or indirectly, considering objective factors including cost, time, available technology, and the state of technological development at the time of processing.

"Blind Signature" means a cryptographic protocol that enables a party to obtain a digital signature on data without revealing the contents of that data to the signer, thereby preserving anonymity while ensuring authenticity.

"Cryptographic Hash Function" means a mathematical function that converts input data of arbitrary size into a fixed-size output, commonly referred to as a hash or digest, in a deterministic, one-way, and collision-resistant manner such that the original data cannot feasibly be reconstructed from the hash.

"Data Augmentation" means techniques used to expand or diversify a dataset for model training by creating modified versions of existing records or generating additional records.

"Data Clean Room" means a secure and controlled environment in which two or more parties analyse combined or linked datasets without either party accessing the other's raw personal data, typically for collaborative analytics.

"Data Commissioner" means the person appointed pursuant to Section 6 of the Act.

"Data Controller" means a natural or legal person, public authority, agency, or other body which, alone or jointly with others, determines the purpose and means of processing personal data.

"Data Masking" means the process of replacing real personal data with realistic but fictitious data in non-production environments, including testing and development environments, to protect personal data from unauthorised exposure.

"Data Processor" means a natural or legal person, public authority, agency, or other body which processes personal data on behalf of the data controller.

"Data Protection Impact Assessment (DPIA)" means a process designed to identify, assess, and mitigate data protection risks arising from processing activities likely to result in a high risk to the rights and freedoms of data subjects, as required under Section 31 of the Act.

"Data Protection Officer (DPO)" means the officer appointed or designated pursuant to Section 24 of the Act.

"Data Subject" means an identified or identifiable natural person who is the subject of personal data.

"Deployer" means an entity that uses or operates a technology, system, or privacy-enhancing technology during its operations and which, in doing so, may act as a data controller or a data processor under the Act.

"Differential Privacy" means a mathematical framework for quantifying and limiting privacy risks to individuals arising from the release of statistical information derived from a dataset by introducing calibrated statistical noise.

"End-to-End Encryption (E2EE)" means an encryption scheme in which data is encrypted at the point of origin and can only be decrypted by the intended recipient, such that no intermediary is able to access the plaintext content.

"Encryption" means the process of converting plaintext data into ciphertext using a cryptographic algorithm and cryptographic key, such that only a party possessing the appropriate key can decrypt the data.

"Epsilon (ε)" means the privacy budget parameter used in differential privacy to quantify the maximum privacy loss permitted by a given query or data release, where a smaller value represents stronger privacy protection.

"Federated Learning" means a machine learning approach in which model training is distributed across multiple devices or data repositories, with participants sharing model updates rather than raw personal data.

"Fully Homomorphic Encryption (FHE)" means a form of encryption that permits arbitrary computations to be performed directly on encrypted data without requiring the data to be decrypted.

"Generalisation" means a privacy-preserving technique in which specific data values are replaced with broader or less precise categories to reduce the risk of re-identification.

"k-Anonymity" means a property of a dataset whereby each record is indistinguishable from at least $k-1$ other records with respect to quasi-identifying attributes.

"l-Diversity" means an extension of k-anonymity that requires diversity in the values of sensitive attributes within each equivalence class.

"Noise Addition" means the deliberate introduction of random statistical perturbations into a dataset or query result to reduce the likelihood of identifying individuals while preserving aggregate analytical utility.

"Office" means the Office of the Data Protection Commissioner established under Section 5 of the Act.

"Personal Data" means any information relating to an identified or identifiable natural person.

"Privacy-Enhancing Technology (PET)" means any technical tool, method, system, or approach that enables the processing, analysis, or sharing of data in a manner that reduces privacy risks, supports compliance with data protection obligations, and preserves the utility of the data.

"Privacy by Design and by Default" means the integration of data protection principles and appropriate technical and organisational measures into the design and default configuration of systems, technologies, and processing activities from the outset.

"Processing" means any operation or set of operations performed on personal data, whether or not by automated means, including collection, recording, organisation, storage, adaptation, retrieval, consultation, use, disclosure, transmission, dissemination, combination, restriction, erasure, or destruction.

"Pseudonymisation" means the processing of personal data in such a manner that it can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and protected by appropriate technical and organisational measures.

"**Regulations"** means the regulations made under Section 71 of the Act and any related subsidiary legislation.

"Secure Aggregation" means a cryptographic protocol that enables a central party to compute an aggregate, including the sum of model updates in federated learning, from inputs supplied by multiple parties without accessing any individual party's input.

"Secure Multi-Party Computation (SMPC)" means a cryptographic protocol that enables multiple parties to jointly compute a function over their combined inputs without revealing their respective inputs to one another.

"Sensitive Personal Data" means data revealing a natural person's race, health status, ethnic or social origin, conscience, belief, genetic data, biometric data, property details, marital status, family details, sex, or sexual orientation.

"Split Learning" means a distributed machine learning technique in which a model is divided between one or more data holders and a central server such that raw personal data remains with the data holder and only intermediate model representations are exchanged.

"Suppression" means the removal of records or values from a dataset to prevent the disclosure of personal data relating to identifiable individuals.

"Synthetic Data" means artificially generated data that is statistically representative of a real dataset but does not correspond to, or identify, any real individual.

"Technology Developer" means an entity that designs, builds, or supplies a technology, system, or privacy-enhancing technology and which, when processing personal data, may act as a data controller or a data processor under the Act.

"t-Closeness" means an extension of l-diversity that requires the distribution of a sensitive attribute within each equivalence class to be sufficiently close to its distribution within the overall dataset.

"Zero-Knowledge Proof (ZKP)" means a cryptographic protocol by which one party proves to another that a particular statement is true without revealing any information beyond the validity of that statement.

1. OFFICE

The Office of the Data Protection Commissioner (Office) is a government agency established to protect the privacy and security of personal data in our increasingly digital world. It is responsible for enforcing data protection laws and policies to safeguard the privacy, dignity, and fundamental rights of individuals. The Office is mandated to oversee the implementation and enforcement of the Data Protection Act, 2019, which regulates the processing of personal data of persons located in Kenya by both private and public sector organisations.

The Office plays a vital role in ensuring that individuals have control over their personal data and that organisations respect their privacy rights. The Office's work involves monitoring and enforcing compliance with data protection regulations, investigating data breaches, and imposing sanctions on entities that violate data protection laws. In addition, the Office is responsible for raising public awareness about data protection issues and educating individuals and organisations on how to protect personal data. With the growing importance of data protection in our digital age, the Office is a critical institution in maintaining trust and confidence in our data-driven society.

The Office is uniquely positioned to facilitate both the government and private sector entities in achieving Government's strategic goals under the "Bottom-Up Economic Transformation Agenda" and, in particular, its digital superhighway initiative. As the digital landscape expands, the need for robust data protection mechanisms becomes paramount. The Office, with its mandate to oversee, regulate, and ensure lawful data processing, plays a pivotal role in this transformation.

Kenya remains at the cutting edge of digital transformation while maintaining stringent data protection standards. The Office serves as a key stakeholder and regulator in guiding the nation's digital superhighway journey by ensuring that as we advance technologically, the rights and privacy of individuals remain safeguarded.

2. INTRODUCTION

2.1. Background

The Data Protection Act establishes a framework of principles and obligations governing the processing of personal data. Those principles are not abstract aspirations: they are legal requirements that shall be given operational effect through appropriate technical and organisational measures. Privacy-Enhancing Technologies are the technical measures through which data controllers and processors translate data protection law into data protection practice.

PETs have been developed and refined over decades of cryptographic and statistical research. Their practical deployment in real-world systems has accelerated rapidly as computational power has increased, as open-source implementations have become widely available, and as regulatory frameworks worldwide have demanded demonstrable privacy controls.

PETs are directly relevant to the challenges of a rapidly digitalising economy: enabling financial inclusion analytics without exposing individual transaction records; supporting national health data analysis without compromising patient privacy; facilitating cross-border data collaboration without violating transfer restrictions; and enabling AI model development without aggregating sensitive training data. PETs are the technical infrastructure of a trustworthy digital economy.

2.2. What are Privacy-Enhancing Technologies?

Privacy-Enhancing Technologies are technical tools, methods, systems, or approaches that enable the processing, analysis, or sharing of data in a manner that reduces or eliminates privacy risks to data subjects. They operate across the data lifecycle at the point of collection, during processing and storage, in analysis and computation, and at the point of sharing or disclosure.

PETs range from foundational cryptographic tools (encryption, hashing) to sophisticated statistical and computational frameworks (differential privacy, secure multi-party computation, federated learning) to data transformation techniques (anonymization, pseudonymization, synthetic data generation).

In practice, organizations often combine more than one PET to build stronger protection for example, using federated learning alongside differential privacy, or pairing pseudonymization with a data clean room.

The table below categorizes PETs by their primary mechanism, and the data protection function each serves

PET Category	Primary Mechanism	Primary Data Protection Function
Data Transformation	Anonymisation, pseudonymisation, generalisation, suppression, data masking	Reduce or eliminate identifiability of data subjects in datasets
Cryptographic	Encryption, hashing, zero-knowledge proofs, blind signatures, secure multi-party computation, homomorphic encryption	Protect data confidentiality and integrity; enable privacy-preserving computation and verification
Statistical / Computational	Differential privacy, noise addition, k-anonymity, l-diversity, t-closeness	Enable statistical analysis and publication of data insights without individual disclosure
Distributed Computation	Federated learning, secure aggregation, split learning	Enable collaborative model training and analysis without centralising personal data
Data Synthesis	Synthetic data generation, data augmentation	Create statistically representative datasets without using real personal data
Controlled Environment	Data clean rooms, trusted execution environments	Enable multi-party data collaboration with access and use controls

2.3. Privacy and Data Protection Concerns Addressed by PETs

PETs address the following data protection concerns that arise in modern data processing environments:

- **Over-collection and retention:** PETs such as differential privacy and anonymisation enable entities to derive analytical value from data without retaining identifiable personal data, directly supporting data minimisation and storage limitation.

- **Breach and unauthorised access:** Encryption ensures that personal data that is accessed without authorisation cannot be read or used, reducing the harm caused by security incidents.
- **Cross-border transfer risk:** PETs such as federated learning and secure multi-party computation enable data collaboration and analysis across jurisdictions without the need to transfer raw personal data, reducing cross-border transfer risk.
- **Inference and re-identification:** Techniques such as differential privacy, k-anonymity, and synthetic data generation reduce the risk that an adversary can re-identify data subjects from published statistical outputs or shared datasets.
- **Verification without disclosure:** Zero-knowledge proofs enable verification of claims (such as age, identity, or qualification) without disclosing the underlying personal data, directly giving effect to data minimisation.
- **Third-party access in collaborative analytics**: Data clean rooms and secure multi-party computation enable organisations to collaborate on data analytics without exposing their raw datasets to each other or to third parties.
- **Insider threat and unauthorised internal access:** Employees, contractors, and system administrators may access personal data beyond the scope of their authorisation, whether inadvertently or deliberately. PETs such as encryption with customer-managed keys, pseudonymisation, and trusted execution environments ensure that personal data is not accessible in identifiable form even to personnel with general system privileges.
- **AI model memorisation and training data exposure:** Machine learning models trained on personal data can memorise individual records and reproduce them in model outputs, exposing personal data through the model itself. PETs such as differential privacy applied to the training process and synthetic training data directly mitigate this risk by bounding the influence of any individual's data on the trained model.
- **Surveillance and profiling through data aggregation:** Combining data from multiple sources each appearing benign in isolation can enable detailed profiling of individuals without their knowledge. This risk is particularly acute in telecommunications and mobile money contexts. PETs such as local differential privacy and pseudonymisation reduce the ability of any entity to aggregate a comprehensive profile of an individual across data sources.
- **Unlawful secondary use and purpose creep**: Personal data collected for a specified purpose may subsequently be used for incompatible purposes, by the original controller or a third party. PETs such as cryptographic access controls, federated learning architectures, and data clean room environments enforce use restrictions technically, preventing data from being extracted or repurposed beyond the agreed terms.

2.4. Scope and Purpose of Guidance Note

This Guidance Note provides data controllers, data processors, technology developers, and deployers with practical guidance on the use of PETs to give effect to data protection obligations under the Data Protection Act and the attendant Regulations. It addresses the principal categories of PETs in current use, their data protection properties and limitations, their relationship to the Act's principles and obligations, and their application in specific sectoral contexts relevant to Kenya.

This Guidance Note applies to all entities, whether public or private, that process personal data of persons located in Kenya and that are considering, deploying, or governing the use of PETs in their data processing operations. It considers:

a) The Data Protection Act, Cap 411C;
b) The Data Protection (General) Regulations, 2021;
c) The Data Protection (Registration of Data Controllers and Data Processors) Regulations, 2021;
d) The Constitution of Kenya, 2010;
e) Other applicable laws and regulations; and
f) International best practices and comparative PET governance frameworks.

3. LEGISLATIVE FRAMEWORK

The adoption and implementation of Privacy-Enhancing Technologies (PETs) is supported by the Data Protection Act, particularly through the obligation to implement privacy by design and by default.

Section 41 of the Act requires data controllers and data processors to implement appropriate technical and organisational measures that are designed to implement the data protection principles effectively and to integrate the necessary safeguards into processing activities, both at the time of determining the means of processing and throughout the processing lifecycle. These measures should ensure that, by default, only personal data that is necessary for each specific purpose of the processing is processed.

The Office considers Privacy-Enhancing Technologies to be among the most effective technical measures for giving practical effect to the requirements of Section 41. The implementation of appropriate PETs assists data controllers and data processors to minimise the collection and disclosure of personal data, reduce privacy risks, enhance data security, and demonstrate compliance with the accountability obligations under the Act. Accordingly, entities should consider the use of appropriate PETs when designing, developing, procuring, deploying, and operating systems or technologies that process personal data.

4. PETS AND DATA PROTECTION PRINCIPLES

Each of the data protection principles under Section 25 of the Act can be given practical effect through the deployment of appropriate PETs. The following table maps the Act's principles to the PETs most directly relevant to their implementation:

Data Protection Principle	**Directly Supporting PETs**
Lawfulness, Fairness, and Transparency	Consent management tools create verifiable records of the lawful basis relied upon for each processing activity and of data subject consent, while cryptographic audit logs ensure those records remain tamper-evident and cannot be altered after the fact. Zero-knowledge proofs that allow an entity to know that a lawful basis for processing exists such as valid consent or a contractual obligation without disclosing the underlying personal data of the data subjects concerned.

	Privacy dashboards and data subject portal tools that provide data subjects with accessible, real-time visibility of the data held about them and the lawful basis for processing.
Purpose Limitation	Data clean rooms and federated learning architectures enforce use restrictions technically, ensuring that participants in a data collaboration can only perform agreed computations and cannot extract data beyond the permitted scope. Cryptographic access controls ensure that personal data can only be accessed or used by systems and personnel with a verified and documented need, preventing its repurposing beyond the original collection purpose. Pseudonymisation limits the ability to link personal data across systems for incompatible purposes, as re-linking requires access to a separately held and controlled key. Tokenisation that replaces direct identifiers with purpose-specific tokens, preventing data from being linked or reused across systems for incompatible purposes.
Data Minimisation	Differential privacy enables statistical analysis and publication of aggregate insights without the need to collect or retain individual-level personal data, directly limiting the volume of personal data processed. Synthetic data replaces real personal data in testing, analytics, and AI training environments, eliminating the need to expose or retain actual personal data for those purposes. Anonymisation gives effect to storage limitation by irreversibly removing identifiers from personal data once it is no longer needed in identifiable form, converting it to non-personal data. Zero-knowledge proofs allow an entity to verify a claim about a data subject such as age, identity, or eligibility without collecting or retaining the underlying personal data that supports that claim. Data masking that replaces real personal data with fictitious values eliminating unnecessary exposure of personal data in development and testing.
Accuracy	Secure aggregation protocols combine personal data from multiple sources while preserving its accuracy, ensuring that the aggregation process itself does not introduce errors or distortion into the resulting dataset.

	Cryptographic integrity checks detect any unauthorised modification of personal data, providing assurance that data has not been altered between collection and use. Audit trails and access logging that create tamper-evident records of data modifications, enabling detection and correction of unauthorised or erroneous changes to personal data.
Storage Limitation	Anonymisation applied at the end of the retention period converts personal data into non-personal data, giving effect to the storage limitation principle without requiring physical deletion of the underlying records. Automated deletion and pseudonymisation workflows triggered by retention schedule events ensure that personal data is not retained in identifiable form beyond the period necessary for the processing purpose. Cryptographic deletion, in which the encryption key for a dataset is securely destroyed, rendering the encrypted personal data permanently inaccessible where technical deletion is not feasible
Integrity and Confidentiality	Encryption at rest and in transit ensures that personal data stored on devices, databases, and media, and transmitted across networks, cannot be read by unauthorised parties even if intercepted or physically accessed. End-to-end encryption ensures that personal data is encrypted at the point of origin and can only be decrypted by the intended recipient, with no intermediate party including the service provider able to access the content. Key management systems govern the generation, storage, rotation, and revocation of cryptographic keys, ensuring that encryption remains effective and that access to protected personal data is controlled and auditable. Secure multi-party computation enables joint processing of personal data by multiple parties without any party accessing the others' raw data, protecting confidentiality throughout the computation. Trusted execution environments provide hardware-level isolation for the processing of personal data, ensuring that even the operating system or cloud infrastructure cannot access data being processed within the secure enclave. Access control and authentication systems ensure that personal data can only be accessed by verified and authorised users, limiting exposure to those with a documented and legitimate need. Hardware security modules (HSMs) for secure generation, storage, and management of cryptographic keys protecting personal data.

Accountability	Cryptographic audit logs provide tamper-evident records of who accessed or processed personal data and when, supporting the controller's ability to demonstrate compliance at any point. DPIA documentation records the PETs selected for each high-risk processing activity, the risks they address, and any residual risks that remain. Privacy-by-design documentation demonstrates that PETs were considered and integrated at the system design stage rather than retrofitted as an afterthought. Periodic PET review records documenting that deployed PETs have been assessed against advances in technology and emerging re-identification risks, and updated where necessary.

The Office emphasises that PETs are technical measures that support compliance: they do not, of themselves, establish compliance. An entity that deploys encryption but has no lawful basis for processing personal data is not compliant with the Act. PETs shall be deployed in conjunction with, and in support of, all applicable data protection principles and obligations, not as a substitute for them.

5. ANONYMISATION AND PSEUDONYMISATION

5.1. Introduction

Anonymisation is the process of irreversibly transforming personal data in such a manner that the data subject can no longer be identified, directly or indirectly, by any person using means that are reasonably likely to be employed. Truly anonymised data is no longer personal data for the purposes of the Act and is therefore not subject to the Act's obligations. Achieving genuine anonymisation is significantly more technically demanding than is commonly assumed, and many datasets that are treated as anonymous in practice remain re-identifiable through linkage with other datasets or through computational inference.

The standard for anonymisation under the Act is objective and absolute: data is anonymised only if re-identification is not reasonably possible given all means likely to be available to any potential adversary, including the data controller and third parties. Entities shall assess the risk of re-identification by reference to the full range of available auxiliary data, linkage techniques, and computational capabilities, not merely the most obvious attack vectors.

5.2 Principal Anonymisation Techniques

Anonymisation is achieved through one or a combination of techniques, depending on the nature of the dataset and the re-identification risks present. Commonly applied techniques include:

- **Generalisation:** Replacing specific values with broader categories (for example, replacing exact postcode with county; replacing precise age with age band). Generalisation reduces precision but may preserve analytical utility.
- **Suppression:** Removing records or values from a dataset that would otherwise allow identification of a data subject, particularly outliers who cannot be generalised without destroying utility.
- **Data masking:** Replacing real personal data with realistic but fictitious values in non-production environments, preventing developers and testers from accessing real personal data.
- **Noise addition:** Adding random statistical perturbation to values in a dataset to prevent accurate inference of individual values, while preserving aggregate statistical properties.
- **k-Anonymity:** Transforming a dataset so that each record is indistinguishable from at least k-1 other records with respect to quasi-identifying attributes, providing a minimum re-identification threshold. k-Anonymity alone may be insufficient where sensitive attributes are not diverse within equivalence classes.
- **l-Diversity and t-Closeness:** Extensions of k-anonymity that additionally require diversity (l-diversity) or distributional closeness (t-closeness) of sensitive attribute values within equivalence classes, addressing attribute disclosure risks not addressed by k-anonymity alone.

5.3 Limitations and Re-identification Risk

Anonymisation techniques shall be assessed against the specific re-identification risks relevant to the dataset and deployment context. The Office notes the following principal re-identification risks:

- **Linkage attacks:** Combining the ostensibly anonymised dataset with publicly available or separately held auxiliary datasets can restore identifiability. High-profile research has demonstrated that individuals can be re-identified in anonymised health records, location datasets, and financial transaction records using publicly available information.
- **Inference attacks:** Statistical inference techniques can extract information about individuals from aggregate query results, even without direct linkage to auxiliary data. Differential privacy is the principal technical defence against inference attacks.
- **Singling out:** Even without knowing an individual's identity, the ability to isolate one individual's record within a dataset constitutes a form of identification risk under data protection law.
- **Data sparsity:** In sparse datasets where individuals have unique combinations of attributes, even heavily generalised records may be effectively unique and therefore identifiable.

Entities that publish or share datasets as anonymised shall document their anonymisation assessment, including:

- the techniques applied;
- the re-identification risk assessment;
- the auxiliary data considered; and
- the conclusion that re-identification is not reasonably possible.

The Office may inspect this documentation in the course of compliance audits.

Example: Health Data Analytics

A Medical Research Agency proposes to share a dataset of patient records with a research consortium for epidemiological analysis. Before sharing, the Agency applies a combination of generalisation (replacing county of residence with region; replacing exact age with five-year age band), suppression (removing records of patients with very rare conditions who would be identifiable through their diagnosis), and an assessment of re-identification risk using publicly available census and health survey data.

The agency documents the assessment, concludes that re-identification is not reasonably possible given the applied techniques and the available auxiliary data, and shares the dataset as anonymised. The Agency reviews the anonymisation assessment if additional auxiliary data becomes available that could affect re-identification risk.

5.4 Pseudonymisation

Pseudonymisation replaces direct identifiers in a dataset such as name, national ID number, or telephone number with a pseudonym (such as a randomly generated token or a cryptographic hash of the identifier), while retaining the ability to re-link records to the original data subject using separately held additional information.

Pseudonymised data remains personal data under the Act, because re-identification remains possible using the additional information. It is therefore subject to the full range of data protection obligations.

Pseudonymisation is nonetheless a valuable risk-reduction measure that the Office actively encourages. It reduces the risk of harm from a data breach (because the pseudonymised dataset alone does not permit identification without the key), supports data minimisation by limiting the circulation of directly identifying data, and can facilitate research, testing, and analytics on sensitive datasets with reduced privacy risk.

5.4.1 Pseudonymisation Techniques

- **Tokenisation:** Replacing a direct identifier with a randomly generated token, with a separately maintained mapping table linking tokens to original identifiers. The mapping table shall be held separately and secured with strict access controls.
- **Cryptographic hashing:** Applying a one-way hash function to an identifier to produce a deterministic but non-reversible pseudonym. Hashing is vulnerable to dictionary attacks where the identifier space is small or predictable (for example, national ID numbers). To mitigate this vulnerability, salted hashing is used, which involves adding a random value before hashing.
- **Encryption-based pseudonymisation:** Encrypting identifiers with a key held separately from the pseudonymised dataset. Re-identification requires possession of the encryption key. The key shall be managed with appropriate security controls.

5.4.2 Key Management for Pseudonymisation

The security of a pseudonymisation scheme depends entirely on the security of the key or mapping table used to re-link pseudonyms to identifiers. Entities shall implement key management practices that ensure:

- the key is held separately from the pseudonymised dataset;
- access to the key is restricted to authorised personnel with a documented need;
- the key is protected by appropriate encryption and access controls; and
- the key is subject to a documented rotation and revocation policy.

Example: Clinical Trial Data

PharmaCo Kenya runs a multi-site clinical trial and shares trial data with external biostatisticians for analysis. The company pseudonymises participant records by replacing national ID numbers and names with randomly generated participant codes, retaining the mapping table in a separate, access-controlled system at the sponsor's headquarters. The biostatisticians receive only the pseudonymised dataset and cannot re-identify participants. The mapping table is accessible only to the trial's data management team and is subject to strict access logging. The pseudonymised data remains personal data; PharmaCo's data processing agreement with the biostatisticians prohibits any attempt at re-identification.

6. ENCRYPTION

Encryption is the foundational technical measure giving effect to the integrity and confidentiality principle. It converts readable data (plaintext) into an unreadable format (ciphertext) using a cryptographic algorithm and key, ensuring that the data can only be read by a party possessing the appropriate decryption key. Encryption is a necessary but not sufficient security measure: it shall be deployed in conjunction with appropriate key management, access control, and other organisational security measures.

The Office requires all entities processing personal data, particularly sensitive personal data to implement encryption as a baseline technical measure. The specific encryption standards and key lengths required will depend on the sensitivity of the data and the threat environment; entities should apply currently recommended standards, noting that cryptographic standards evolve as computational capabilities advance.

6.1. Encryption at Rest and in Transit

Entities shall implement encryption of personal data both at rest (in storage) and in transit (during transmission over networks). Encryption at rest protects personal data stored on servers, databases, laptops, mobile devices, and removable media from unauthorised access in the event of device loss, theft, or unauthorised physical access. Encryption in transit protects personal data from interception as it is transmitted between systems, devices, and parties over networks, including the internet.

- Encryption at rest:
  - Full-disk encryption for endpoint devices;
  - database-level encryption or field-level encryption for personal data in databases;
  - file-level encryption for personal data in cloud storage;
  - encryption of backup media.
- Encryption in transit:
  - Transport Layer Security (TLS) for web applications and API communications;
  - Secure Shell (SSH) for administrative access;

- encrypted messaging protocols for communications containing personal data;
- VPN for remote access to systems containing personal data.

6.2. End-to-End Encryption

End-to-end encryption (E2EE) ensures that data is encrypted at the point of origin and can only be decrypted by the intended recipient, with no intermediate party including the service provider able to access the plaintext content. E2EE is particularly important for communications services, health platforms, legal services, and any application where the confidentiality of the content of communications or data is paramount and where service provider access to content is not required for the service to function.

Entities deploying E2EE shall disclose the use of E2EE in their privacy notices and shall assess whether E2EE is compatible with their obligations to respond to data subject rights requests and to implement retention and deletion obligations in respect of encrypted content.

6.3. Key Management

The security of an encryption scheme is only as strong as the security of the key management practices supporting it. Entities shall implement key management practices that address:

- key generation (using cryptographically secure random number generators and appropriate key lengths);
- key storage (hardware security modules or equivalent secure key storage);
- key access controls (limiting access to encryption keys to authorised personnel and systems with a documented need);
- key rotation (periodic rotation of encryption keys and re-encryption of data);
- key revocation (ability to revoke compromised keys and re-encrypt affected data); and
- key escrow and recovery (documented procedures for key recovery in the event of loss, without enabling unauthorised access).

Example: Mobile Health Application

AfyaApp, a mobile health application that collects patient symptom data and medication records, implements encryption at rest using AES-256 for data stored on users' devices and in the cloud database, and TLS 1.3 for all data transmitted between the app and the server. The company implements customer-managed encryption keys in its cloud environment, ensuring that the cloud provider cannot access plaintext health data. The privacy notice discloses the encryption standards applied. The DPO reviews the key management policy annually.

7. DIFFERENTIAL PRIVACY

Differential privacy (DP) is a method of sharing statistics or insights from a dataset while protecting the privacy of each individual in it. It works by making sure the final results would look nearly the same whether one person's data is included or left out. This means no one can easily tell if a specific person was part of the dataset.

To do this, a small amount of carefully controlled randomness is added to the results. For example, a government may publish population trends, or a hospital may share disease statistics without

exposing any single person's information. In simple terms, it helps organisations learn from the crowd without revealing details about the individual.

Differential privacy is operationalised by adding calibrated statistical noise to query outputs or to the dataset itself. The amount of noise is governed by the privacy budget parameter epsilon (ε): a smaller ε means more noise and stronger privacy protection but reduced statistical accuracy; a larger ε means less noise, higher accuracy, but weaker privacy protection. Entities shall set ε by reference to the sensitivity of the data and the acceptable trade-off between privacy protection and analytical utility.

7.1. Applications of Differential Privacy

Differential privacy is applicable wherever an entity wishes to:

- publish statistical outputs such as aggregate counts, averages, or distributions derived from a dataset of personal data, without enabling inference about any individual's data;
- release query interfaces such as statistical APIs or dashboards over personal data, while preventing the accumulation of information about individuals through repeated or crafted queries;
- train machine learning models on sensitive personal data, adding noise to the training process to prevent the model from memorising individual records; or
- share data with third parties for analytics purposes, where raw personal data cannot or should not be shared.

Differential privacy is deployed at scale by major technology companies and government statistical agencies.

7.2. Local and Global Differential Privacy

Global differential privacy: Noise is applied to the aggregate output of a query on a centralised dataset. The data controller holds the raw dataset and applies differential privacy at the point of output release. Global DP provides strong privacy guarantees with lower noise (and therefore higher analytical utility) for a given ε.

Local differential privacy: Noise is applied to each individual's data on their device before transmission to the data controller. The data controller never receives raw personal data, only already-perturbed values. Local DP is appropriate where the data controller itself should not have access to individual-level raw data. It requires more noise (and therefore produces lower analytical utility) than global DP for a given ε.

7.3. Limitations of Differential Privacy

Differential privacy has important limitations that entities shall understand:

- Privacy budget composition: Each query against a differentially private dataset consumes privacy budget. Repeated queries accumulate privacy loss; entities shall manage and bound the total privacy budget consumed across all queries on a given dataset.
- Utility trade-off: Stronger privacy protection (smaller ε) requires more noise, which reduces the accuracy and utility of published statistics. Entities shall determine an ε that is appropriate for the specific analytical purpose and acceptable privacy risk.

- Not a universal solution: Differential privacy is designed for statistical outputs, not for protecting individual records in operational databases. It is not a substitute for encryption in transactional systems.

Example: National Health Statistics

The Kenya National Bureau of Statistics (KNBS) publishes county-level health outcome statistics derived from a national patient registry. To protect patient privacy particularly in counties with small populations where aggregate statistics could enable identification of individuals with rare conditions KNBS applies differential privacy with a calibrated ε of 1.0 to all published county-level statistics, adding Laplace-distributed noise to counts and averages before publication. The statistical bulletin discloses that differential privacy has been applied and provides guidance on interpreting perturbed statistics.

8. FEDERATED LEARNING AND DISTRIBUTED COMPUTATION

Federated learning (FL) is a machine learning approach in which a model is trained across multiple devices or data silos without centralising the underlying training data. Instead of sending personal data to a central server, each participant trains the model locally on their own data and sends only model updates gradients or parameter adjustments to a central aggregator, which combines them to update the global model. Raw personal data never leaves the participant's environment.

Federated learning directly supports data minimisation and purpose limitation by enabling model training without requiring the aggregation of personal data at a central location. It reduces cross-border transfer risks where participants are in different jurisdictions. It is particularly relevant for health, financial services, and telecommunications use cases where personal data is sensitive and/or legally or contractually constrained from being shared.

8.1. Applications of Federated Learning

- Health analytics: Hospitals and health facilities can collaboratively train diagnostic AI models on patient data without sharing patient records across institutions or with commercial model developers.
- Credit scoring: Mobile network operators and financial institutions can collaboratively train credit risk models using their respective customer data without sharing raw transaction records, protecting customer privacy and competitive confidentiality.
- Fraud detection: Banks and payment processors can collaboratively train fraud detection models on their respective transaction datasets without centralising transaction data, reducing both privacy risk and data governance complexity.
- Agricultural analytics: Agri-tech platforms serving smallholder farmers can train yield prediction and disease detection models on farmer data held locally on devices or at cooperative level, without aggregating individual farm data centrally.

8.2. Privacy Risks and Mitigations in Federated Learning

While federated learning significantly reduces the privacy risks associated with centralised data collection, it is not inherently privacy-preserving. The following risks shall be addressed:

- Gradient inversion attacks: Sophisticated adversaries can reconstruct training data from the model gradients shared by participants. Secure aggregation, a cryptographic protocol in which gradients are aggregated without the aggregator seeing any individual participant's gradient and differential privacy applied to the gradients (differentially private SGD) mitigate this risk.
- Model inversion and membership inference: The trained global model may memorise information about training data, enabling adversaries to infer whether a particular individual's data was in the training set. Differential privacy applied to the training process bounds this risk.
- Poisoning attacks: Malicious participants can corrupt the global model by submitting adversarially crafted gradient updates. Robust aggregation protocols and anomaly detection in gradient updates mitigate this risk.

Entities deploying federated learning for processing involving personal data shall conduct a DPIA that addresses these specific risks and the technical mitigations adopted. Where federated learning involves participants in multiple jurisdictions, entities shall assess cross-border transfer obligations in respect of any data including model updates that flows between jurisdictions.

Example: Federated Learning for Cross-Bank Fraud Detection

Three Kenyan commercial banks propose to collaborate on training an AI fraud detection model using their combined transaction data. Rather than pooling transaction records at a central location which would create data protection, competitive confidentiality, and cross-institution governance challenges the banks deploy a federated learning architecture in which each bank trains the model locally on its own transaction data and shares only gradient updates with a secure aggregation server. No raw transaction data leaves any bank's environment. The aggregation server uses secure aggregation to combine gradients without any party seeing another bank's individual gradients. Differential privacy is applied to the aggregated gradients before the global model is updated. The banks conduct a joint DPIA addressing gradient inversion and membership inference risks.

9. SECURE MULTI-PARTY COMPUTATION

Secure Multi-Party Computation (SMPC) is a cryptographic framework that enables two or more parties to jointly compute a function over their combined inputs without any party revealing its input to the other parties or to the computation infrastructure. SMPC provides a cryptographic guarantee that each party learns only the output of the agreed computation, and nothing about any other party's input beyond what can be inferred from the output itself.

SMPC is particularly powerful for enabling privacy-preserving analytics and collaboration between organisations that hold complementary datasets but cannot or will not share raw personal data: competitors who wish to compute industry benchmarks; hospitals that wish to compute aggregate disease statistics without sharing patient records; or government agencies that wish to compute population statistics by linking datasets without creating a centralised linked database.

9.1. Applications of SMPC

- Privacy-preserving record linkage: Two or more organisations can identify records that correspond to the same individual across their respective datasets without revealing the records themselves to each other. For example, a tax authority and a social protection agency could identify individuals who appear in both datasets to prevent fraud, without sharing their respective databases.
- Aggregate statistics and benchmarking: Competing entities can compute industry-wide statistics (average salary, market share, fraud rates) without revealing their individual data to each other or to a trusted third party.
- Joint AI model training: Multiple organisations can train a joint AI model on their combined data without any organisation's raw data leaving its environment, providing stronger privacy guarantees than federated learning in certain threat models.
- Secure voting and auditing: Cryptographic voting protocols using SMPC enable the tallying of votes without any single party having access to individual vote choices, applicable to governance and audit processes.

9.2. Limitations and Practical Considerations

SMPC protocols are computationally intensive and impose significant performance overhead relative to plaintext computation, limiting their practical applicability to scenarios where the computation is sufficiently valuable to justify the cost. The communication overhead between parties in an SMPC protocol scales with the complexity of the computation and the number of parties. Practical SMPC deployments typically require careful protocol selection, pre-computation of cryptographic material, and dedicated infrastructure.

Entities considering SMPC should engage specialist cryptographic expertise in protocol design and implementation. Incorrectly implemented SMPC protocols can provide weaker guarantees than intended. The DPIA for any processing involving SMPC should document the specific protocol used, the security model assumed, and the trust assumptions made about each party's computational capabilities.

Example: Privacy-Preserving Salary Benchmarking

Seven Kenyan insurance companies wish to compute industry-wide average salaries by grade to benchmark their compensation structures against competitors, without any company revealing its own salary data to the others. The companies implement an SMPC protocol in which each company encrypts its salary data using secret shares and distributes the shares among all participants. Each participant computes a partial result using its share, and the partial results are combined to produce the industry average, without any participant learning any other company's individual salary data. Only the aggregate industry average is revealed to all participants.

10. ZERO-KNOWLEDGE PROOFS

A zero-knowledge proof (ZKP) is a cryptographic protocol in which a prover demonstrates to a verifier that a statement is true without revealing any information beyond the truth of the statement itself. ZKPs enable verification without disclosure: a party can prove that they satisfy a condition, such as being above a certain age, holding a valid credential, or having a sufficient account balance without revealing the underlying personal data that constitutes the evidence.

ZKPs can support the data minimisation principle by enabling verification of specific attributes without disclosing the underlying personal data: where a verification objective can be achieved through a zero-knowledge proof, in many implementations the verifier may only receive confirmation that a condition has been satisfied. Entities should nonetheless assess whether metadata, transaction logs, identifiers, or other ancillary information generated by the proof process constitutes personal data. ZKPs are applicable wherever a relying party needs to verify a claim about a data subject without requiring access to the underlying personal data.

10.1. Applications of Zero-Knowledge Proofs

- Age and identity verification: A data subject can prove that they are above a minimum age, or that they are a citizen of Kenya, without disclosing their date of birth, national ID number, or other personal details to the verifying party. The verifier receives only a cryptographic proof that the condition is satisfied.
- Credential and qualification verification: A data subject can prove that they hold a valid professional licence, academic qualification, or government-issued credential without disclosing the credential document or the issuing authority's records to the verifying party.
- Financial and solvency verification: A financial institution can verify that a counterparty has sufficient funds or meets a creditworthiness threshold without the counterparty disclosing their account balance or credit score.
- Digital identity systems: National digital identity frameworks can use ZKPs to enable citizens to prove specific attributes of their identity, such as residency, tax status, or benefits eligibility to service providers without disclosing the full contents of their identity record. This directly minimises the data collected by service providers and reduces the risk of identity data aggregation.
- Regulatory compliance: Regulated entities can prove compliance with a regulatory requirement, such as meeting a capital adequacy ratio or holding a required licence to a regulator without disclosing confidential business data beyond what is necessary for the compliance verification.

10.2. Limitations and Implementation Considerations for ZKPs

ZKPs are an enabling technology rather than a guarantee of compliance, and entities deploying them shall take account of the following limitations:

- Setup and implementation risks: Some ZKP schemes rely on a trusted setup; compromise of the setup parameters can undermine the validity of proofs, and implementation flaws in proving or verification code can create vulnerabilities that are difficult to detect.
- Key and credential management: The security of ZKP-based verification depends on the secure issuance, storage, rotation, and revocation of the underlying credentials and cryptographic keys.

- Metadata and ancillary data: ZKP transactions may generate metadata, transaction logs, and identifiers that may themselves constitute personal data and shall be assessed and protected accordingly.
- Computational cost and maturity: ZKP schemes vary in maturity and computational cost, and entities shall assess the suitability of a scheme for the scale and latency requirements of the deployment.

Entities considering ZKP deployments should engage specialist cryptographic expertise, and the DPIA for a ZKP deployment shall address these limitations and the associated mitigation measures.

Example: Age Verification for Digital Financial Services

A Kenyan digital wallet provider shall verify that new customers are above the minimum age of 18 before onboarding. Currently, the provider collects and retains a copy of each customer's national ID card. Under a ZKP-based age verification system, the customer's device generates a zero-knowledge proof derived from the customer's verified digital ID credential that they are over 18, without disclosing their date of birth or any other personal data to the wallet provider. The provider verifies the proof cryptographically and completes onboarding. The provider receives only the verification outcome and the associated technical metadata necessary to complete the transaction, supporting the data minimisation principle.

11. HOMOMORPHIC ENCRYPTION

Homomorphic encryption (HE) is a form of encryption that permits specified computations to be performed directly on encrypted data (ciphertext), producing an encrypted result that, when decrypted, is identical to the result of performing the same computation on the unencrypted plaintext. Fully Homomorphic Encryption (FHE) permits arbitrary computations on ciphertext; Partially Homomorphic Encryption (PHE) permits only specific operations (such as addition or multiplication but not both).

Homomorphic encryption enables a powerful privacy model: a data owner can send encrypted personal data to a cloud service provider or analytics platform, which performs the required computation on the ciphertext without ever decrypting the data, and returns the encrypted result to the data owner, who decrypts it. Where properly implemented, the service provider processes ciphertext and is unable to access plaintext personal data without the corresponding decryption key.

Homomorphic encryption operates within defined threat assumptions and trust boundaries. It protects personal data during processing, but it does not eliminate risks arising at the endpoints where data is encrypted and results are decrypted, from malicious insiders with access to decryption keys, from side-channel attacks, or from misconfiguration and key management failures. Entities shall document the threat model and trust boundaries of any homomorphic encryption deployment, including which parties hold decryption keys and which components process plaintext.

11.1. Applications of Homomorphic Encryption

- Encrypted cloud analytics: An organisation can upload encrypted personal data to a cloud analytics platform, which performs statistical queries or machine learning inference on the ciphertext, without the cloud provider accessing the underlying personal data. Highly relevant for health analytics, financial modelling, and census data processing.
- Secure medical diagnosis: A patient's health data can be sent encrypted to an AI diagnostic service, which runs a diagnostic model on the ciphertext and returns an encrypted diagnosis result to the patient or their physician, without the diagnostic service provider accessing the patient's health data.
- Privacy-preserving credit scoring: A credit bureau can perform credit assessment computations on encrypted financial data provided by a bank, returning an encrypted credit score to the bank, without the credit bureau being able to access the underlying account data.

11.2. Current Limitations

Fully Homomorphic Encryption is computationally intensive orders of magnitude slower than equivalent plaintext computation which currently limits its practical applicability to scenarios where the privacy benefit justifies the performance cost. Practical FHE deployments typically require careful circuit design to minimise computational depth, specialised hardware acceleration, and significant engineering investment. The field is advancing rapidly: performance improvements and hardware acceleration are making FHE increasingly practical for specific high-value use cases.

Key management is the primary determinant of the security effectiveness of homomorphic encryption. Entities shall apply the key management requirements set out in this Guidance Note to homomorphic encryption deployments, including secure key generation, storage in hardware security modules or equivalent, access controls, rotation, and revocation, and shall manage the

cryptographic lifecycle of deployed schemes, including migration away from deprecated parameters and algorithms.

Entities considering homomorphic encryption should engage specialist cryptographic expertise. The DPIA for HE-based processing should document the specific HE schemes used, the computation permitted, and the security parameters selected.

12.SYNTHETIC DATA

Synthetic data is artificially generated data that is statistically representative of a real dataset, preserving the statistical distributions, correlations, and structural properties of the original data but is intended not to correspond directly to any real individual. Synthetic data is generated by training generative models on real personal data and then using those models to generate new, fictitious data records.

Synthetic data enables entities to share, publish, test on, and develop AI systems using data that preserves the analytical utility of real data without processing real personal data. Synthetic data can support data minimisation with respect to the downstream activity (testing, sharing, AI training) where it reduces reliance on identifiable personal data, subject to validation of the associated privacy risks.

12.1. Applications of Synthetic Data

- Software testing and development: Developers and testers can work with realistic, structurally representative data without accessing real personal data, eliminating the risk of developer exposure to sensitive data.
- AI and machine learning training: AI models can be trained on synthetic data instead of or in addition to real personal data, reducing the volume of personal data collected and retained for training and the associated privacy risks.
- Data sharing and publication: Synthetic datasets can be shared with researchers, analysts, and partners as a privacy-preserving substitute for the real dataset, enabling data collaboration without disclosing personal data.
- Regulatory reporting and analytics: Entities can generate synthetic versions of regulatory reporting datasets for internal testing and audit preparation without using live personal data.

12.2. Limitations and Data Protection Status of Synthetic Data

Synthetic data is not automatically anonymous. The generative model used to produce synthetic data is trained on real personal data, creating two distinct risks:

- Memorisation: Generative models can memorise specific records from the training data and reproduce them in synthetic outputs, particularly for rare or outlier records. An entity shall assess whether the synthetic dataset contains records that correspond to real individuals before treating it as non-personal.
- Linkage and inference: Even if no individual synthetic record corresponds to a real person, the statistical properties of the synthetic dataset may enable an adversary to infer characteristics of the real training data or of specific individuals whose data was disproportionately influential in the training process.

Entities using synthetic data should assess whether a DPIA is required based on the nature, scope, context, and purposes of the processing and the associated risks, and shall assess the synthetic dataset before treating it as non-personal data exempt from the Act. The assessment should

evaluate memorisation risk (using membership inference testing) and statistical disclosure risk. Where the assessment confirms that synthetic data poses no reasonable re-identification risk, it may be assessed as anonymised data where the risk of re-identification is demonstrably remote, taking into account all reasonably likely means of re-identification. Where such confirmation is not possible, the synthetic data should be treated as pseudonymised personal data and handled accordingly.

Example: Synthetic Health Records for AI Development

HealthTech Kenya develops an AI-based patient triage system. The company first establishes a lawful basis for processing the original training data used to generate the synthetic dataset, conducts a DPIA, and implements data protection controls for that training dataset. It then generates a synthetic patient record dataset using a variational autoencoder trained on a smaller real dataset obtained under appropriate consent. Before using the synthetic dataset for AI training, the company conducts a membership inference test to assess whether individual real patient records can be reconstructed from the synthetic data, and a statistical comparison to confirm that the synthetic dataset faithfully represents the statistical properties of the real data. The membership inference test confirms no material memorisation risk. The company treats the synthetic dataset as anonymised data not subject to the Act for purposes of the AI training activity.

13. DATA CLEAN ROOMS

A data clean room is a secure, controlled, and privacy-governed environment in which two or more parties can jointly analyse combined or linked datasets without direct access to the other party's raw personal data, subject to the technical and governance controls implemented within the clean room environment. Clean rooms enforce privacy controls, including minimum aggregation thresholds, output restrictions, and audit logging that prevent any participant from extracting personal data about the other's data subjects while still enabling agreed analytical computations.

Data clean rooms are widely used for privacy-preserving advertising measurement, financial services analytics, healthcare research, and government data collaboration. They enable data-driven collaboration between organisations that hold complementary datasets but cannot share raw personal data for legal, contractual, or competitive reasons.

13.1. Data Protection Properties of Clean Rooms

- Data isolation: Each party's raw data remains within that party's environment; only query results subject to privacy controls are released.
- Minimum aggregation thresholds: Clean room environments enforce minimum group sizes for query outputs (for example, suppressing results where fewer than ten individuals would be represented), preventing identification through small-group queries.
- Output restrictions: Agreed restrictions on the type and granularity of outputs that can be extracted from the clean room, preventing the reconstruction of raw data from iterative queries.
- Audit logging: All queries and outputs are logged, enabling review of the analytical activities conducted in the clean room and detection of attempts to exceed agreed use restrictions.

- Contractual governance: Clean room use is governed by a data collaboration agreement that specifies the permitted use cases, output restrictions, parties' data protection responsibilities, and audit rights.
- Role-based access control and segregation of duties: Access to the clean room environment is restricted through role-based access controls, and duties are segregated so that no single individual can both configure the privacy controls and extract outputs.
- Independent audit and assurance: The clean room environment and its privacy controls are subject to periodic independent audit and assurance, with findings documented and remediated.

13.2. Obligations When Using Data Clean Rooms

Entities using data clean rooms for processing personal data shall:

- identify the lawful basis for each party's contribution of personal data to the clean room environment;
- document the data controller and data processor relationships among the parties involved in the clean room, noting that in some configurations each party may be a controller of their own dataset and a joint controller of the combined analytics output;
- conduct a DPIA before deploying a clean room for high-risk processing activities, including advertising targeting, credit analytics, or health research;
- ensure that the privacy controls implemented by the clean room environment (minimum thresholds, output restrictions, audit logging) are documented and verified;
- ensure that the data collaboration agreement governing clean room use addresses all data protection obligations applicable to each party; and
- assess cross-border transfer obligations where the clean room infrastructure or any party's data is located outside Kenya.

Example: Retailer-Bank Data Collaboration

A Kenyan retail chain and a commercial bank wish to measure the effectiveness of a joint loyalty programme by analysing purchase behaviour in relation to credit card usage. Neither party is willing to share its customer data with the other. They deploy a data clean room in which each party loads a pseudonymised dataset. The clean room computes agreed aggregate analytics, such as the proportion of cardholders who made a qualifying purchase in each campaign period and releases only aggregate outputs that exceed a minimum group threshold of 50 individuals. No individual-level records are accessible to either party. The parties sign a data collaboration agreement documenting the permitted use cases, the privacy controls applied, and their respective data protection responsibilities.

14. PETS IN SPECIFIC SECTORAL CONTEXTS

PETs are applicable across all sectors that process personal data. The following guidance addresses PET deployment in the sectors of greatest data protection significance in Kenya's digital economy.

14.1. Financial Services and Mobile Money

Kenya's financial services sector including commercial banks, microfinance institutions, mobile money platforms , digital lenders, and insurance companies processes some of the largest volumes of personal financial data in the country. PETs directly applicable to this sector include:

- Differential privacy for fraud analytics and risk reporting: Enabling aggregate fraud statistics and risk reporting without exposing individual transaction data.
- Federated learning for credit scoring and fraud detection: Enabling collaborative model training across financial institutions without sharing customer transaction records.
- SMPC for regulatory reporting: Enabling financial sector aggregates to be computed across institutions and reported to regulators without any institution disclosing individual account data to others.
- Tokenisation and pseudonymisation of payment card data: Reducing the exposure of payment card data in transaction processing systems.
- Encryption of transaction records at rest and in transit: Protecting transaction data from unauthorised access in the event of breach.
- Zero-knowledge proofs for KYC verification: Enabling identity and solvency verification for financial services onboarding without requiring the collection and retention of full identity documents and financial records.

14.2. Healthcare and Digital Health

Kenya's healthcare sector, including public hospitals and clinics, private health providers, health insurers, and digital health platforms processes highly sensitive patient health data at significant scale. PETs applicable to this sector include:

- Pseudonymisation of patient records for research and analytics: Enabling clinical research and population health analytics on pseudonymised records, with the re-identification key held separately under strict access controls.
- Differential privacy for public health statistics: Enabling publication of county and national health statistics without enabling identification of patients with rare conditions in small geographic areas.
- Federated learning for diagnostic AI: Enabling collaborative training of diagnostic AI models across hospital networks without sharing patient records.
- Homomorphic encryption for remote health analytics: Enabling cloud-based health analytics on encrypted patient data without the analytics platform accessing raw health records.
- Synthetic health data for AI development and testing: Replacing real patient records with statistically representative synthetic data in AI training and software testing environments.
- Encryption and access controls for electronic health records: Protecting patient data in electronic health record systems from unauthorised access, with role-based access controls limiting access to the minimum data required for each clinical or administrative function.
- PETs supporting health research ethics and secondary use: Enabling ethically approved health research and the secondary use of health data for research and public health analytics, through anonymisation, pseudonymisation, and synthetic data, in accordance with applicable research ethics approvals and the Act.

14.3. Public Sector and Government

Government agencies and public institutions at national and county level process vast volumes of citizen personal data across functions including civil registration, taxation, social protection, education, land administration, and public health. PETs applicable to the public sector context include:

- Federated and privacy-preserving data linkage for programme delivery: Enabling benefits eligibility assessment and service delivery by linking records across agencies without creating centralised databases of linked citizen data. The Estonia X-Road model may serve as an illustrative example of privacy-preserving data exchange architecture.
- Differential privacy for census and survey publications: Protecting individual respondents in national statistics publications, particularly at sub-county geographic levels where small populations create identification risk.
- Zero-knowledge proofs for citizen identity verification: Enabling citizens to prove specific identity attributes (residency, tax status, eligibility for a public programme) to service delivery agencies without those agencies retaining full identity records.
- Encryption of government databases: Protecting citizen data in government databases from breach, with key management practices that ensure accountability for access.
- Pseudonymisation for inter-agency data sharing: Pseudonymising citizen records before sharing between agencies for agreed analytical or administrative purposes, limiting the downstream identifiability of shared data.

14.4. Telecommunications and Digital Platforms

Mobile network operators, internet service providers, and digital platforms process location, communication metadata, and usage data about large proportions of Kenya's population. PETs applicable to this sector include:

- Differential privacy for network analytics and usage statistics: Enabling aggregate network performance and usage reporting without exposing individual user activity.
- Local differential privacy for user behaviour analytics: Enabling collection of app usage and behaviour telemetry from user devices with noise applied on-device before transmission, so that the platform never receives un-perturbed individual usage data.
- Pseudonymisation of metadata records: Pseudonymising call detail records and location data in analytics systems, limiting access to identified data to compliance and lawful interception functions subject to strict controls.
- Encryption and end-to-end encryption for communications: Protecting message content and call data from interception and unauthorised access.
- Location data re-identification risks: Location and movement data is among the most re-identifiable categories of personal data, and even coarse or sampled location traces can single out an individual. Entities shall apply heightened PETs, including aggregation, differential privacy, and strict retention limits, before location data is analysed or shared.

15. PETS AND DATA SUBJECT RIGHTS

PETs interact with data subject rights in important ways. The deployment of PETs can both facilitate and complicate the fulfilment of data subject rights. The following guidance addresses the principal interactions.

15.1. Right of Access and Pseudonymisation

Where personal data has been pseudonymised, data subjects retain the right of access to the personal data held about them under the Act. The pseudonymisation of data does not extinguish the data subject's rights in respect of that data. Entities shall implement mechanisms to re-link a data subject's access request to their pseudonymised records, retrieve those records, and supply the data subject with their personal data in a comprehensible format. Access to the re-linking key for the purpose of fulfilling an access request shall be subject to appropriate access controls and logged.

15.2. Right to Erasure and Anonymisation

Where a data subject requests erasure and the conditions for erasure under the Act are met, anonymisation of the data subject's records irreversibly removing all information that could identify the data subject may constitute compliance with the erasure obligation where technical erasure is not feasible (for example, in aggregate statistical datasets or trained AI models). Entities relying on anonymisation as a substitute for erasure shall confirm that the anonymisation is irreversible and complete and shall document the anonymisation as the erasure mechanism applied.

Where a data subject requests erasure of data that has been used to train an AI model, entities shall assess whether the model memorises the individual's data (through membership inference testing) and implement machine unlearning or model retraining where memorisation is identified.

15.3. Right to Rectification and Differential Privacy

Where personal data that has been processed through differential privacy mechanisms is subsequently found to be inaccurate, the data subject's right to rectification applies to the underlying personal data held by the controller, not to published statistical outputs. Entities shall correct inaccurate personal data in their source records and, where the inaccuracy has materially affected published statistics, assess whether a correction or retraction of the affected statistics is warranted.

15.4. Right to Data Portability and Encryption

The right to data portability requires personal data to be provided in a structured, commonly used, and machine-readable format. Where personal data is stored in encrypted form with customer-managed keys, entities shall be able to provide data subjects with a decrypted, portable copy of their personal data in a standard format. The encryption of data does not relieve the entity of its portability obligations.

16. OBLIGATIONS OF DATA CONTROLLERS AND DATA PROCESSORS REGARDING PETS

16.1. Privacy by Design: PET Adoption as a Statutory Obligation

Section 41 of the Act requires data controllers to implement appropriate technical and organisational measures to give effect to data protection principles. The adoption of appropriate PETs is a direct expression of this obligation. The Office expects data controllers to:

- assess the PETs available for each significant processing activity at the system design stage, document the assessment, and select the most appropriate PET or combination of PETs given the nature of the processing, the sensitivity of the data, and the state of available technology;
- implement selected PETs as core system features rather than as retrospective add-ons;
- review PET deployments periodically to ensure they remain appropriate as technology advances and as the risk profile of the processing activity changes; and
- document PET deployment decisions in the records of processing activities maintained under the Act.

The fact that a PET is technically complex, commercially costly, or operationally inconvenient does not relieve a data controller of the obligation to implement it where it is appropriate and reasonably available. The Office will consider the state of available technology and the cost of implementation in assessing whether an entity has met the privacy by design obligation, but expects entities to invest appropriately in PETs as a cost of responsible data processing.

16.2. DPIA and PET Documentation

The DPIA required under Section 31 of the Act for high-risk processing shall include a description of the technical and organisational measures proposed to mitigate identified risks. PETs adopted as risk mitigations shall be documented in the DPIA, including: the specific PET selected; the risk or risks it addresses; the technical implementation; and any residual risks that the PET does not fully address. The adoption of appropriate PETs may reduce the assessed risk level of a processing activity and may obviate the need for prior consultation with the Office.

16.3. Data Processor Agreements and PET Requirements

Where a data processor processes personal data on behalf of a controller, the data processing agreement required by the Data Protection (General) Regulations, 2021 shall specify the technical and organisational measures the processor is required to implement. For high-risk or sensitive data processing, the agreement should specify the required PETs: for example, requiring a cloud analytics provider to implement differential privacy in published outputs; requiring an AI developer to use federated learning or synthetic data rather than raw personal data; or requiring an identity verification service to implement ZKP-based verification where technically feasible.

16.4. Data Sharing Agreements and PET Requirements

Where personal data is shared with another data controller, including within a data collaboration or clean room arrangement, the parties shall execute a data sharing agreement that specifies the PETs and other technical and organisational measures to be applied to the shared data, the purpose and lawful basis of the sharing, and the responsibilities of each party for giving effect to data subject rights. Before sharing, entities shall assess whether the sharing objective can be achieved using a PET that avoids the disclosure of identifiable personal data, such as pseudonymisation, secure multi-party computation, or a data clean room, and shall document that assessment.

16.5. Obligation to Comply with Data Protection Principles and PET Requirements

Data controllers and data processors shall process personal data in accordance with the data protection principles under Section 25 of the Act in all processing activities. PETs shall be selected and configured to give effect to each applicable principle, including lawfulness, fairness and transparency, purpose limitation, data minimisation, accuracy, storage limitation, and integrity and confidentiality, as set out in this Guidance Note. The deployment of a PET does not displace the obligation to comply with each principle, and entities shall verify, on an ongoing basis, that PET deployments continue to support compliance with the principles.

16.6. Staff Competency and Training

Data controllers and processors shall ensure that staff responsible for data protection governance including DPOs, data protection leads, IT architects, and procurement officers have sufficient understanding of available PETs to assess their appropriate deployment in the entity's processing activities. This does not require all relevant staff to be technical cryptography experts but does require a working understanding of the data protection properties, limitations, and deployment considerations of the principal PETs addressed in this Guidance Note. The Office provides capacity building on PETs through its training programme and stakeholder engagement activities.

16.7. Registration

All entities that process personal data of persons located in Kenya including entities deploying PETs in connection with such processing shall be registered with the Office as data controllers or data processors in accordance with the Data Protection (Registration of Data Controllers and Data Processors) Regulations, 2021. The Office has published a Guidance Note on Registration of Data Controllers and Data Processors accessible at www.odpc.go.ke.

17. ENFORCEMENT AND COMPLAINTS

The Office is mandated to oversee the implementation and enforcement of the Act. The Office is empowered to receive and investigate complaints, conduct inspections and audits of data controllers and data processors, issue enforcement and penalty notices, impose administrative fines, and make referrals for criminal prosecution where applicable, including in respect of entities that fail to implement appropriate technical and organisational measures to give effect to data protection principles.

The Office will treat the failure to implement appropriate PETs particularly where a high-risk processing activity involves sensitive personal data and technically and commercially feasible PETs are available as evidence of non-compliance with the privacy by design obligation under Section 41 of the Act and the integrity and confidentiality principle.

Data subjects who believe that an entity has failed to implement appropriate technical measures to protect their personal data, or who have concerns about the validity of an entity's anonymisation or pseudonymisation practices, may lodge a complaint with the Office. Complaints may also relate to:

- failure to implement encryption of sensitive personal data at rest or in transit;
- representation that personal data has been anonymised where re-identification remains reasonably possible;
- failure to implement a DPIA that addresses technical risk mitigations for high-risk processing;

- failure to specify required PETs in data sharing agreements and in data processing agreements with processors engaged in high-risk processing;
- processing of personal data contrary to the data protection principles in circumstances where appropriate PETs would have supported compliance;
- failure to give effect to other data subject rights, including access, rectification, objection, and portability, in systems where PETs have been deployed; or
- failure to give effect to a data subject's erasure right in a system where anonymisation or cryptographic deletion is technically feasible.

The Office encourages entities to engage proactively with the Office on PET adoption questions, including through participation in ODPC-facilitated stakeholder engagement and capacity building programmes. Proactive engagement and investment in appropriate PETs will be considered in the Office's assessment of an entity's compliance posture.

To lodge a complaint or seek further guidance, contact the Office at:

Office of the Data Protection Commissioner

Britam Towers, 12th Floor, Hospital Road, Upperhill, Nairobi

P.O. Box 30920-00100, Nairobi, Kenya

Email: complaint@odpc.go.ke (for lodging complaints) | info@odpc.go.ke / enquiries@odpc.go.ke (for enquiries and further guidance)

Telephone: 0207801800 | 0796954269 / 0752896867

Website: www.odpc.go.ke

ANNEX 1: PET SELECTION GUIDE

The following guide assists entities in identifying appropriate PETs for common data processing objectives. It is intended as a starting point for PET selection; entities should consult specialist technical expertise when designing and implementing PET deployments.

Processing Objective	Recommended PET(s)	Key Considerations
Protect personal data from breach in storage	Encryption at rest	Implement with key management system; address backup and replica encryption
Protect personal data during network transmission	TLS 1.3 or equivalent; E2EE for communications content	Ensure certificate management; E2EE where service provider access to content is not required
Enable statistical analysis and publication without individual disclosure	Differential privacy; k-anonymity / l-diversity / t-closeness for tabular data where formal privacy guarantees are not require	Differential privacy is the preferred standard for high-sensitivity data and open publication, providing formally verifiable guarantees. k-anonymity and its extensions are weaker, heuristic protections vulnerable to composition and attribute disclosure attacks; use only for lower-sensitivity datasets under controlled access. Set ε by reference to data sensitivity and manage privacy budget across queries.
Train AI models without centralising personal data	Federated learning with secure aggregation and differential privacy	Address gradient inversion risk; conduct DPIA; assess cross-border transfer if participants in multiple jurisdictions
Share data between organisations for analytics without disclosing raw data	Secure multi-party computation; data clean rooms; pseudonymisation	SMPC for high-security requirements; clean rooms for commercial analytics; assess legal basis for sharing
Replace real personal data in testing and development environments	Synthetic data generation; data masking	Assess memorisation risk in synthetic data; membership inference testing before treating as non-personal
Verify identity or attributes without collecting underlying personal data	Zero-knowledge proofs	Requires ZKP-enabled identity infrastructure; disclose to data subjects in privacy notice
Analyse encrypted data without decrypting it	Homomorphic encryption	High computational cost; assess feasibility for specific computation; engage specialist expertise
Enable inter-agency government data exchange with privacy controls	Federated architecture; pseudonymisation; differential privacy for aggregate outputs	Reference Estonia X-Road as architecture benchmark; implement tamper-evident access logging

Reduce re-identification risk in shared research datasets	Anonymisation (generalisation + suppression + noise); differential privacy; synthetic data	Conduct formal re-identification risk assessment before releasing as anonymous; document assessment
Enable end-to-end audit of data access without exposing content	Cryptographic audit logs; trusted execution environments	Implement immutable logging; access controls on audit log system
Minimise health data processed by diagnostic AI platforms	Federated learning; homomorphic encryption; synthetic training data	Layer PETs: federated learning for training; HE for inference; synthetic for development

ANNEX 2: COMPLIANCE CHECKLIST FOR PET DEPLOYMENTS

The following checklist assists entities in self-assessing compliance with data protection obligations in connection with PET deployments. It does not constitute legal advice and does not substitute for a full DPIA or specialist technical review.

Assessment Area	What to Check For	Yes	No	Recommended Action if No
Governance and Accountability	Has the organisation designated a DPO or equivalent responsible for PET oversight, with a documented policy governing PET selection, deployment, and review?	☐	☐	Designate a DPO or equivalent and develop a documented PET governance policy covering selection, deployment, and periodic review.
Privacy by Design	Are PETs considered and documented at the system design stage rather than retrofitted after deployment?	☐	☐	Embed PET consideration into the system design process and update system architecture documentation to evidence PET integration at design stage.
DPIA Integration	Is a DPIA conducted before high-risk processing, documenting the specific PETs selected, risks addressed, and residual risks remaining?	☐	☐	Conduct a DPIA for all high-risk processing activities and ensure PET selection, risk mitigation, and residual risks are documented within each DPIA.
Anonymisation and Pseudonymisation	Is a re-identification risk assessment conducted before data is treated as anonymous, and are pseudonymisation keys stored separately with restricted access?	☐	☐	Implement a formal re-identification risk assessment procedure and establish separate, access-controlled storage for pseudonymisation keys.
Encryption	Is personal data encrypted at rest and in transit using currently recommended standards, supported by a documented key management policy?	☐	☐	Implement encryption at rest and in transit across all systems processing personal data and develop a key management policy covering generation, storage, rotation, and revocation.
Differential Privacy	Where differential privacy is applied, is the epsilon (ε) value documented, justified, and the total privacy budget tracked across all queries?	☐	☐	Document and formally justify the epsilon value applied by reference to data sensitivity and analytical purpose, and implement privacy budget tracking across all queries on each dataset.
Federated Learning	Is raw personal data confirmed to remain within each participant's environment, with gradient inversion and membership inference risks addressed in the DPIA?	☐	☐	Verify that the federated learning architecture prevents raw data from leaving participant environments and update the DPIA to address gradient inversion, membership inference, and poisoning attack risks.

Secure Multi-Party Computation	Has specialist cryptographic expertise been engaged, and is the specific protocol, security model, and trust assumptions documented in the DPIA?	☐	☐	Engage specialist cryptographic expertise to review the SMPC protocol and document the specific protocol used, security model assumed, and trust assumptions in the DPIA.
Zero-Knowledge Proofs	Where ZKPs are used, is it confirmed that no personal data beyond the proof outcome is collected, and is this disclosed in the privacy notice?	☐	☐	Review ZKP implementation to confirm data minimisation is achieved and update the privacy notice to disclose the use of zero-knowledge proofs in verification processes.
Synthetic Data	Has a residual re-identification risk assessment been conducted on synthetic datasets before sharing or publication?	☐	☐	Conduct and document a residual re-identification risk assessment on all synthetic datasets prior to any sharing or publication activity.
Data Clean Rooms and TEEs	Are use restrictions enforced technically and contractually, and are hardware and software attestation mechanisms documented where TEEs are used?	☐	☐	Implement technical use controls within the clean room environment, include use restriction clauses in all data collaboration agreements, and document TEE attestation mechanisms.
Vendor and Third-Party PET Deployments	Are vendor PET technical specifications documented in contracts, with a right to audit and independent verification of vendor claims?	☐	☐	Update vendor contracts to include PET technical specifications, right to audit provisions, and requirements for independent certification or verification of PET implementations.
Data Subject Rights	Are PET deployments assessed for compatibility with the organisation's ability to respond to data subject rights requests including access, correction, and erasure?	☐	☐	Conduct a compatibility assessment of all PET deployments against data subject rights obligations and implement technical or procedural measures to ensure rights can be fulfilled.
Review and Continuous Improvement	Are PET deployments reviewed periodically against advances in technology and emerging attack methods, with a documented process for updating inadequate PETs?	☐	☐	Establish a periodic PET review schedule, assign responsibility to the DPO, and document the process for updating or replacing PETs found to be inadequate against current threat models.

Entities with questions about PET selection, deployment, or compliance assessment are encouraged to contact the Office at the address below or to consult the ODPC's capacity building programme.